# THE EFFECT OF INTERCONNECT CONFIGURATION ON DELAY IN COMPUTERS USING HIGH SPEED CIRCUITS

Dr. E. A. Wilson
Honeywell Information Systems
P. O. Box 6000 B-96
Phoenix, Az. 85005

## INTRODUCTION

As the design of computers moves from utilizing TTL logic circuits to high speed, non-saturating circuits such as HCML (Honeywell's Current Mode Logic), the simplistic method of delay estimation during the design phase by just adding the gate delays is no longer adequate. This is because the delay associated with interconnection can actually exceed the circuit delay (i.e. gate times of one nanosecond or less) in critical timing paths. Of course, using LSI chips which have many serial logic gates per chip eliminates the media delay for internal connections. Honeywell's micropackaging technology (Figure 1),

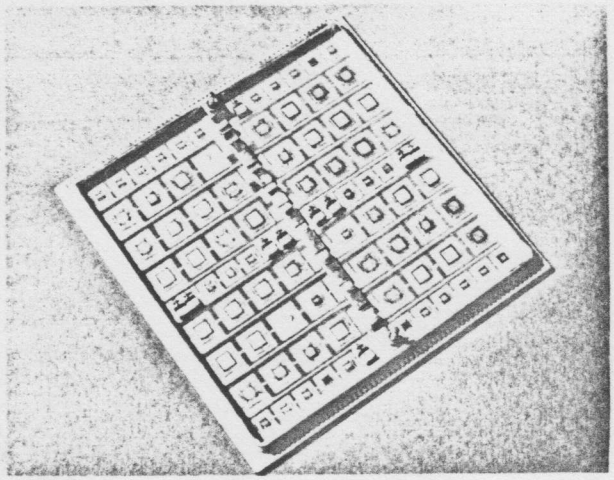


FIGURE 1

which places up to 110 chips on an 80mm by 80mm alumina substrate, significantly reduces the run lengths as compared to DIPs mounted on a printed circuit board.

However, between micropackages the media delay still exists and calculation of the delay becomes more than just multiplying the line length by the propagation speed. In addition to the media propagation speed itself, the geometry of the interconnect (branch points, connectors, various media in the same signal path) has an effect on signal propagation with high speed edges; and the loading on the driving gate varies with the number of driven gates, affecting the rise time of the line voltage. Hence, the resultant delay is truly an interconnect configuration delay, not just a media delay.

The sample problems presented in this paper represent several practical interconnect configurations which will be encountered in large, high speed computer designs. Such effects as turn-on/turn-off/turn-on of the load due to reflected waves and load delay dependence on wave edge deterioration which result from the interconnect configuration are demonstrated by these problems.

Because the clock (cycle) time is considerably faster for a high speed machine (than for a TTL machine), these calculations must be very accurate in order to meet performance goals. The simulation model used for the sample problems is briefly presented in the appendix to this paper.

## MODELING PARAMETERS AND OUTPUT PLOTS

The various media used in the overall packaging which employs micropackages are shown in Figure 2.

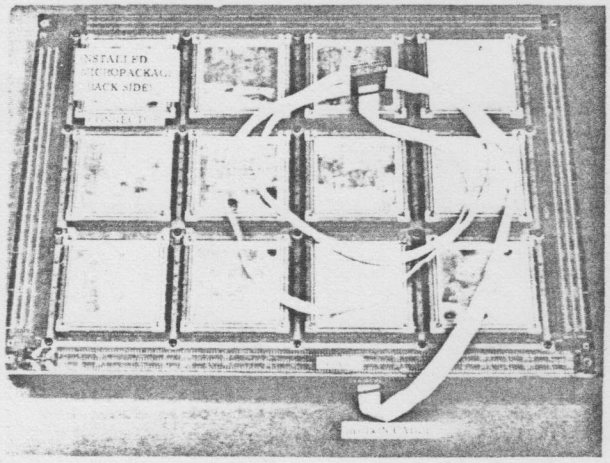


FIGURE 2

They are the micropackage, the connector (for mounting the micropackage to the board), the board, and the ribbon cable which interconnects boards. The table below gives the impedence and propagation speed for each media. The connector and ribbon cable characteristics are essentially non-variable because of their geometry and manufacturing methods while the micropackage and board manufacturing processes inherently produce parts which vary and must be kept within a range by proper quality control procedures. One of the sample problems will demonstrate how these variations affect the interconnect configuration delay.

| MEDIUM | IMPEDANCE ohms | PROPAGATION SPEED nanoseconds/mm |
|---|---|---|
| MICROPACKAGE (nominal) | 34.7 | .01128 |
| (fast) | 44.1 | .01076 |
| (slow) | 28.9 | .01176 |
| BOARD (nominal) | 75 | .0072 |
| (fast) | 83 | .0064 |
| (slow) | 67 | .008 |
| CONNECTOR | 31.6 | .01264 |
| RIBBON CABLE | 75 | .0045 |

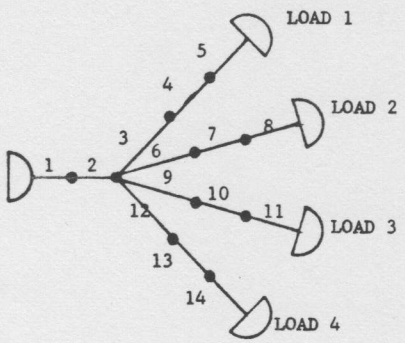
Before presenting the problems and their results, an explanation of the output plots will aid in interpreting the results. As is explained in the appendix, the loads are simulated by a simple HCML gate. Except where noted otherwise, the output plots are the waveforms of the output of the driven gate (or load). The model provides either the load input or output waveforms, but the output waveform is the more significant since it is used to determine the interconnect configuration delay. This is done by noting the time when the output of the load crosses the threshold or switching voltage (.26 volts) and subtracting a constant which had been determined from a gate driving a load with zero interconnection length. The reason for using the load output voltage is that the interconnect affects load turn-on time, hence the load input voltage is not adequate to determine the total effect of the interconnect configuration on delay.

In actual design practice, the result of running the simulation program is just a report back to the user of the interconnect configuration delay for each load. The waveforms are only obtained by a special request from the user if he needs additional details in order to optimize and/or reduce delay of a particular path. The execution speed of the program is so fast that it is practical to simulate the thousands of interconnects during the design phase of a large computer. The design data base automatically provides the input for the simulation program and the calculated delay times are then included as a part of the overall data base for use by the design engineers.

STAR CONFIGURATION

This problem will be developed from a simple, impractical (in the sense that it would not appear in a real design) star to a more realistic interconnect which may loosely be described as a star. Only nominal media parameters will be used in this problem since the intent is to show how various media, variations in load factors and actual geometry affect the delay.

The simple star is shown in Figure 3 and for the first simulation, all of the lines were treated as board lines instead of using the connector or micropackage parameters. Then the micropackage and connector parameters were used for the appropriate segments and the load output plots are shown in Figure 4 along with the driver output voltage for the case of all three media in the problem. Because all signal paths are the same length and all load factors equal, only one load output plot exists for each case. Of note is the fact that when all media are considered, the delay is 3 nanoseconds longer than for the all board lines case. This is important because if the difference in propagation speed between board lines and connector or micropackage lines is multiplied by the appropriate line lengths, only .48 nanoseconds can




| LINE | LENGTH (MM) | MEDIUM |
|---|---|---|
| 1 | 25 | MICROPACKAGE |
| 2 | 25 | CONNECTOR |
| 3 | 775 | BOARD |
| 4 | 25 | CONNECTOR |
| 5 | 25 | MICROPACKAGE |
| 6 | 775 | BOARD |
| 7 | 25 | CONNECTOR |
| 8 | 25 | MICROPACKAGE |
| 9 | 775 | BOARD |
| 10 | 25 | CONNECTOR |
| 11 | 25 | MICROPACKAGE |
| 12 | 775 | BOARD |
| 13 | 25 | CONNECTOR |
| 14 | 25 | MICROPACKAGE |

FIGURE 3

be accounted for in the 3 nanosecond difference. The bulk of the difference is therefore due to the variation of line characteristics in the signal paths. This can be noted by the dip in the source (driver) voltage.

The next case also uses Figure 3 except that the top load has a load factor of 2.5 (for example, two "high" current gates and a "low" current gate driven in parallel on the same chip or adjacent chips) and the other three loads only have a load factor of .5. Note that the total load seen by the source is still 4, the same as the previous case. The output plots are in Figure 5 and show a 1.5 nanosecond difference in delay time. However, the possibly surprising result is that the gates on the more heavily loaded line turn on before the half loads. This is due to the capacitance of the larger loading making the top line less sensitive to the dip in the source voltage which the lightly loaded lines track closer.

A true equal line length star would not likely be found in a real design since board routing and micropackage placement would preclude such an ideal case. The quasi star in Figure 6 is more representative of a real interconnect. The board line lengths are chosen to relate to the previous star configuration. The average distance to the four loads is the same as the previous equal board line lengths. Likewise, the total loading (four) is the same. In Figure 7 only the output of loads 1 and 2 and the source are shown for the sake of clarity. Load 1 turns on 10.5 nanoseconds sooner than for the equal line length case yet the difference in distance of 275 mm only accounts for the signal reaching the load 2 nanoseconds sooner. The reason for the faster turn-on can be seen in the source voltage. It reaches a higher initial plateau which is due to the

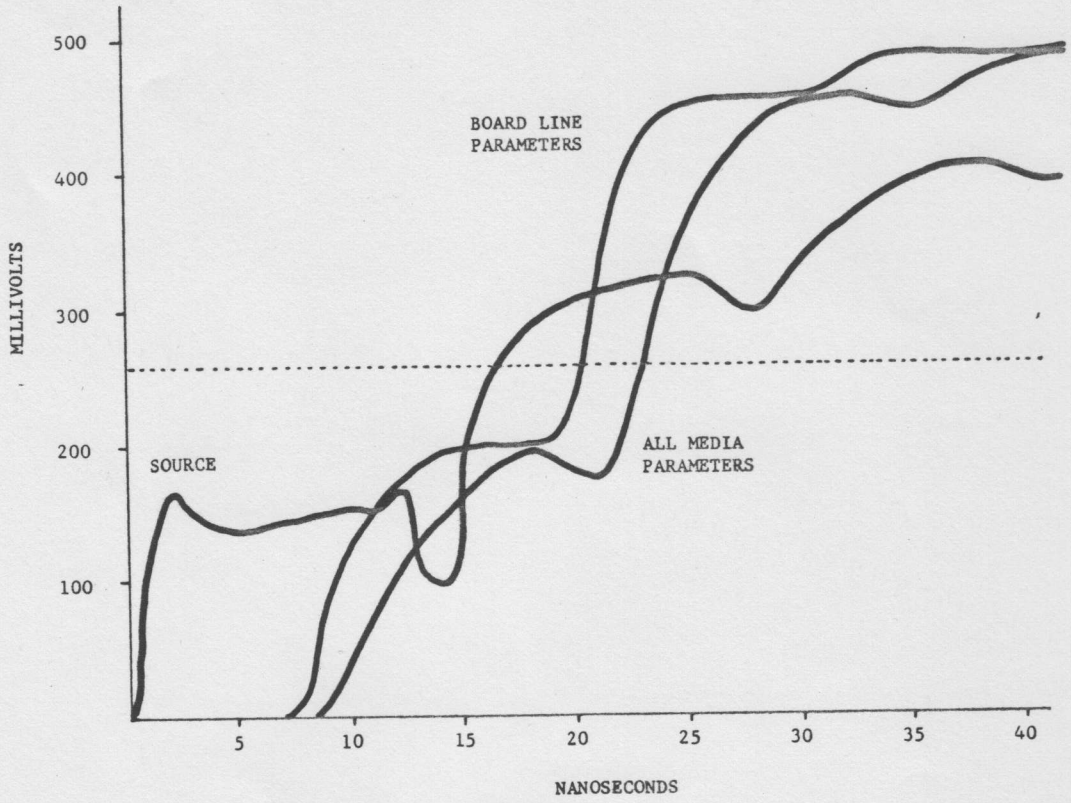



FIGURE 4

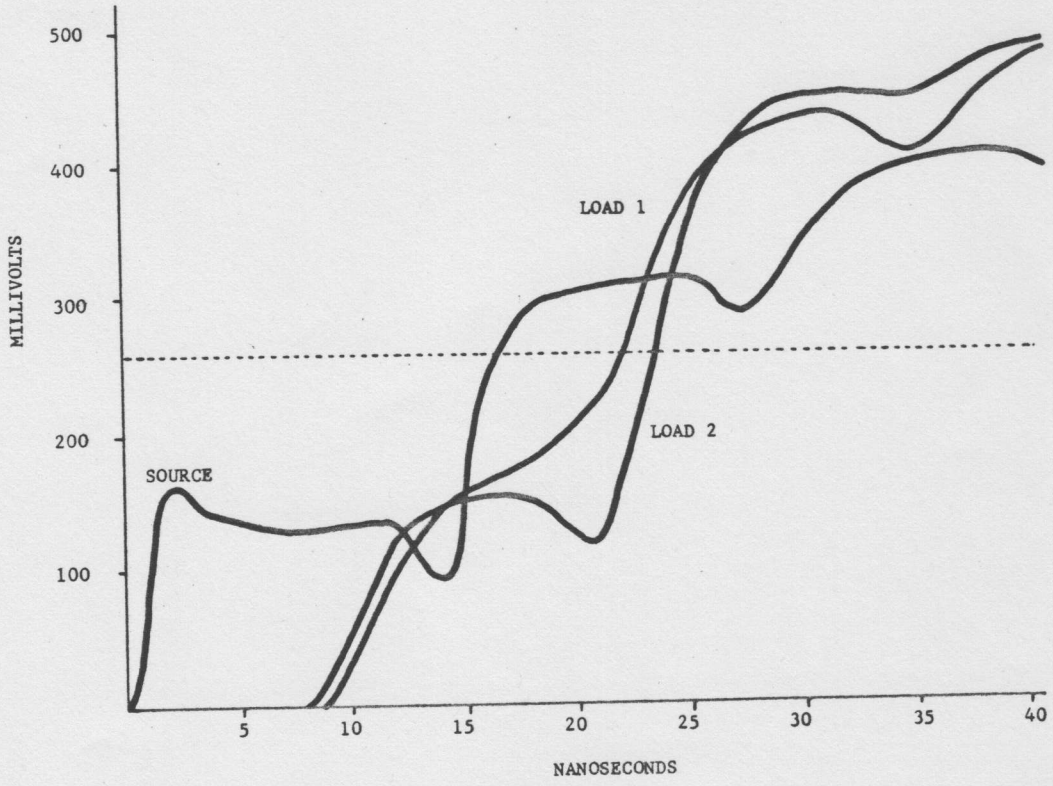
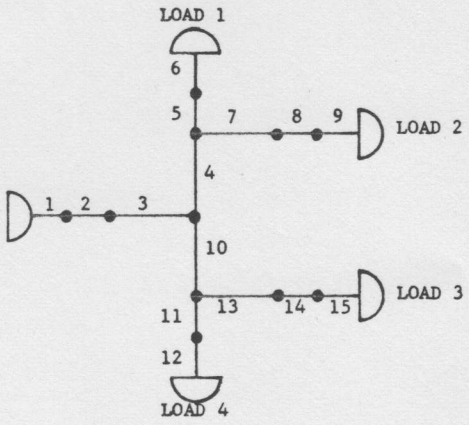
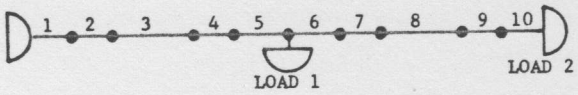



FIGURE 5




| LINE | LENGTH (MM) | MEDIUM |
|---|---|---|
| 1 | 25 | MICROPACKAGE |
| 2 | 25 | CONNECTOR |
| 3 | 50 | BOARD |
| 4 | 450 | BOARD |
| 5 | 25 | CONNECTOR |
| 6 | 25 | MICROPACKAGE |
| 7 | 150 | BOARD |
| 8 | 25 | CONNECTOR |
| 9 | 25 | MICROPACKAGE |
| 10 | 850 | BOARD |
| 11 | 25 | CONNECTOR |
| 12 | 25 | MICROPACKAGE |
| 13 | 150 | BOARD |
| 14 | 25 | CONNECTOR |
| 15 | 25 | MICROPACKAGE |

FIGURE 6




| LINE | LENGTH (MM) | MEDIUM |
|---|---|---|
| 1 | 25 | MICROPACKAGE |
| 2 | 25 | CONNECTOR |
| 3 | 300 | BOARD |
| 4 | 25 | CONNECTOR |
| 5 | 50 | MICROPACKAGE |
| 6 | 50 | MICROPACKAGE |
| 7 | 25 | CONNECTOR |
| 8 | 300 | BOARD |
| 9 | 25 | CONNECTOR |
| 10 | 25 | MICROPACKAGE |

FIGURE 8

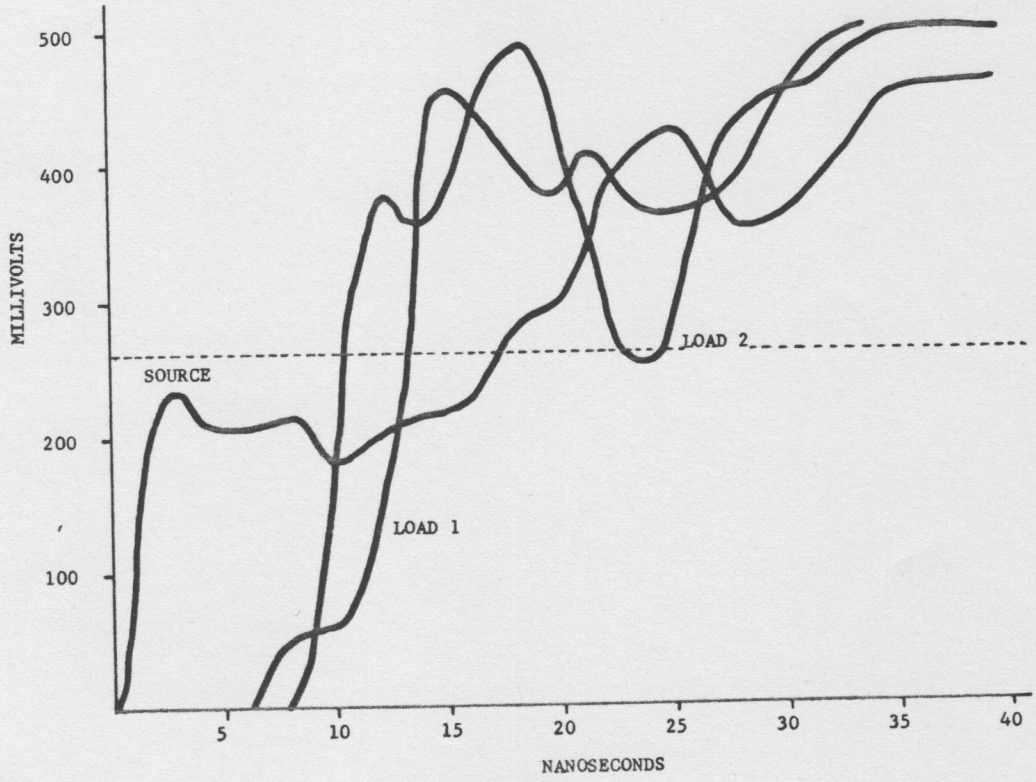



FIGURE 7

t branch point having only two instead of four nches. The waveform of load 2 is also noteworthy. Although it turns on before load 1, it turns off at 23 nanoseconds and on again at 24 nanoseconds. Hence, 24 nanoseconds must be taken as the true turn-on time since the load is itself a driver for the next step in the logic path and must be stable before turn-on of its loads can be assured.

DAISY CHAIN CONFIGURATION

As in the star problem, this will be developed from a simplistic problem to a realistic problem. The interconnect configuration is shown in Figure 8. For the first case all lines are given board parameters and then the real parameters are used. The comparison of load 1 output waveforms for the two cases is shown in Figure 9. Again, the difference in propagation speed for the appropriate lines does not account for the 5 nanosecond difference in turn-on time.

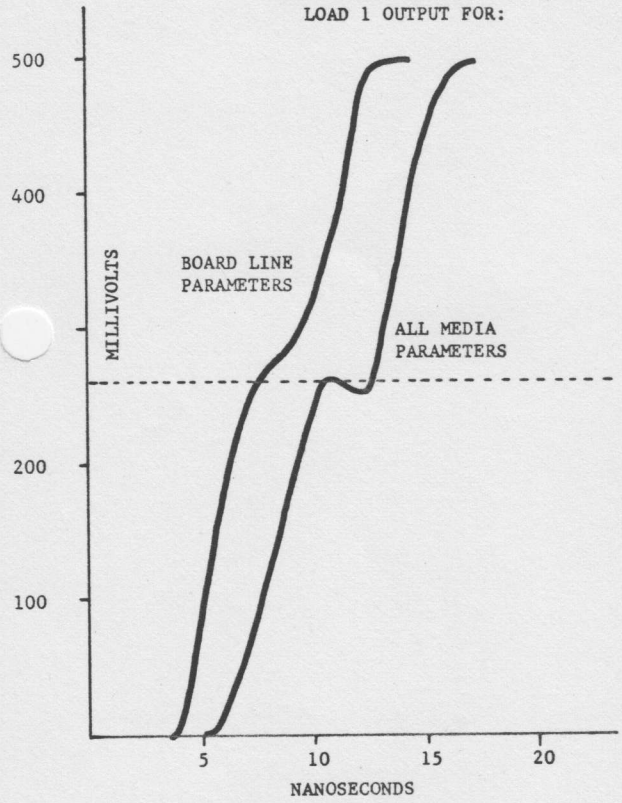


FIGURE 9

The effect of the interconnect configuration on the switching delay through the load gate can be seen in Figure 10 which shows the source voltage, input to load, and load output. The nominal switching delay through the gate for a ramp input voltage is only about one nanosecond. However, the waveform actually presented to the gate by the interconnect is not a ramp, hence the delay through the load gate (time between input reaching threshold voltage and output reaching threshold voltage) is 4 nanoseconds.

By the above cases, the need for accurate model- of the various portions of the interconnect instead of just using propagation speed multiplied by

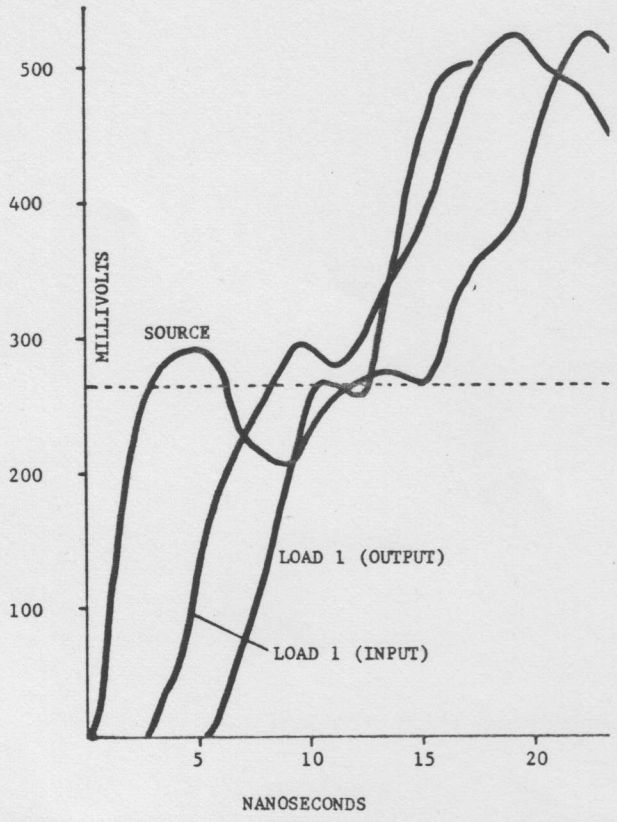


FIGURE 10

-line length is demonstrated. The need to examine the effect of variations in media parameters due to fabrication and process changes is demonstrated by Figure 11. The four combinations of minimum and maximum parameters for the micropackage and board lines were used to produce the four output waveforms for load 1. The turn-on/turn-off/turn-on phenomenon is most severe for slow board and fast micropackage lines while it is non-existant for the other cases. Intuitively, one might think that the slow board and fast micropackage combination would produce a smoother curve than the fast board and slow micropackae combination since the parameters are closer to each other for the former and more mismatched in the latter. However, the reverse is true and demonstrates the need for simulation instead of intuition. The variation in parameters varies the turn-on time by as much as 4 nanoseconds. For this reason, the standard simulation procedure mentioned earlier uses the fast and slow as well as the nominal parameters and the slowest delay at each load calculated from the simulations for the combinations of parameters is saved in the data base as the design delay. This allows a high confidence that the total computer design will meet the performance goals even though media variations (within prescribed limits, of course) occur throughout the build life of the product.

MULTIPLE BRANCH POINTS (and modification of design as a result of waveform examination)

The interconnect shown in Figure 12 is representative of a source in a micropackage on one board driving multiple loads on a second board and a single load such as a latch on a third board. Obviously, such an interconnect configuration could not be

1 - FAST BOARD, FAST MICROPACKAGE
2 - FAST BOARD, SLOW MICROPACKAGE
3 - SLOW BOARD, FAST MICROPACKAGE
4 - SLOW BOARD, SLOW MICROPACKAGE

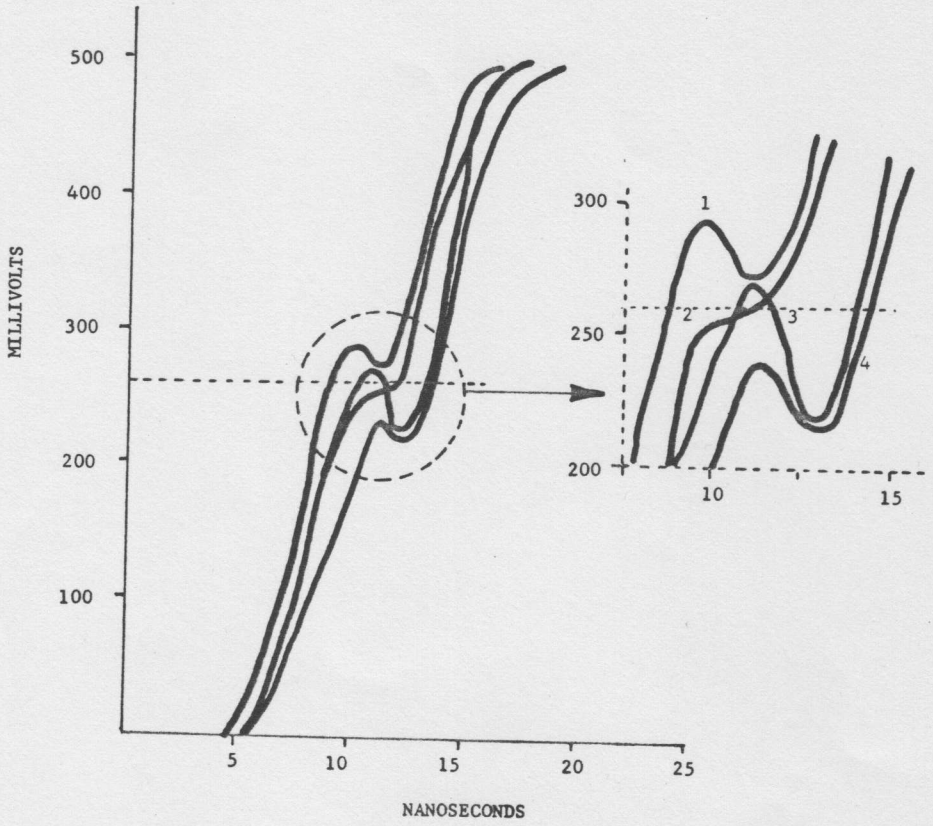



FIGURE 11

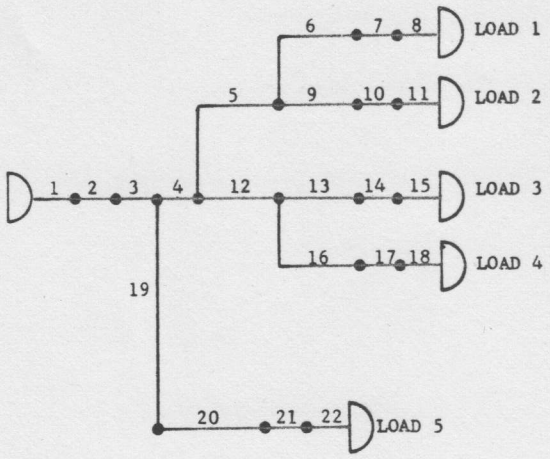
classified as either a star or daisy chain yet it is more realistic than the ideal star. Only the nominal line parameters are used for this problem.

The waveforms of the source and loads 3 and 5 are shown in Figure 13. The respective turn-on times of loads 3 and 5 are 28 and 34 nanoseconds. If this were a situation where load three is in a critical path, the delay could well be excessive. Now consider the modification to the design shown in Figure 14. Load 5 is brought back to the first board and used to drive load 6 which is in the same position that load 5 previously occupied on the third board. The turn-on time for load 6 can be determined by matching the threshold points of the load 5 output and the source for the secondary interconnect simulation, then superpositioning the load 6 output waveform plot on the waveform plots of the primary interconnect. This is done in Figure 15. The result is that by adding the gate (load 5 moved to the first board) the turn-on time on the third board has only been increased from 34 to 35 nanoseconds while the critical load turn-on time has been reduced from 28 to 17 nanoseconds, a significant reduction for a high speed computer running with a fast clock time.

Obviously, such a savings in delay time would have been indeterminate without a simulation program which could adequately model the interconnects shown in Figures 12 and 14.

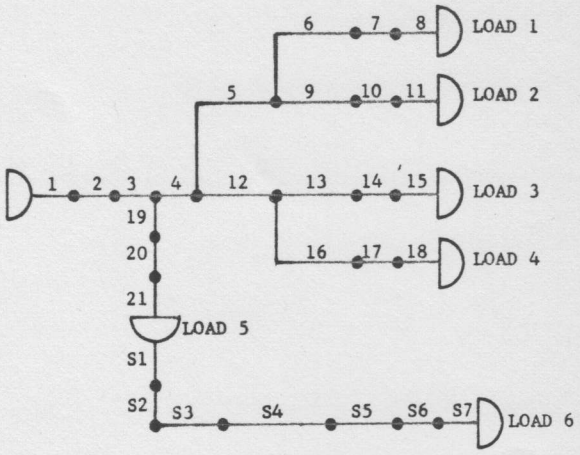
## CONCLUSION

The interconnection of logic gates, and hence the packaging of the total machine, can no longer be left to the end of the design phase of a new computer. It must be considered and factored into the design at the onset of actual logic design, and even sooner if possible. This is because of the impact of the interconnect configuration on delay and machine performance. Also, adequate design tools are required in order to correctly determine how the interconnect configurations will affect the signal propagation. Simple hand calculations can no longer do the job.




| LINE | LENGTH (MM) | MEDIUM |
|---|---|---|
| 1 | 25 | MICROPACKAGE |
| 2 | 25 | CONNECTOR |
| 3 | 250 | BOARD |
| 4 | 750 | RIBBON CABLE |
| 5 | 375 | BOARD |
| 6 | 250 | BOARD |
| 7 | 25 | CONNECTOR |
| 8 | 25 | MICROPACKAGE |
| 9 | 250 | BOARD |
| 10 | 25 | CONNECTOR |
| 11 | 25 | MICROPACKAGE |
| 12 | 125 | BOARD |
| 13 | 375 | BOARD |
| 14 | 25 | CONNECTOR |
| 15 | 25 | MICROPACKAGE |
| 16 | 50 | BOARD |
| 17 | 25 | CONNECTOR |
| 18 | 25 | MICROPACKAGE |
| 19 | 625 | RIBBON CABLE |
| 20 | 625 | BOARD |
| 21 | 25 | CONNECTOR |
| 22 | 25 | MICROPACKAGE |

FIGURE 12




| LINE | LENGTH (MM) | MEDIUM |
|---|---|---|
| 1 | 25 | MICROPACKAGE |
| 2 | 25 | CONNECTOR |
| 3 | 250 | BOARD |
| 4 | 750 | RIBBON CABLE |
| 5 | 375 | BOARD |
| 6 | 250 | BOARD |
| 7 | 25 | CONNECTOR |
| 8 | 25 | MICROPACKAGE |
| 9 | 250 | BOARD |
| 10 | 25 | CONNECTOR |
| 11 | 25 | MICROPACKAGE |
| 12 | 125 | BOARD |
| 13 | 375 | BOARD |
| 14 | 25 | CONNECTOR |
| 15 | 25 | MICROPACKAGE |
| 16 | 50 | BOARD |
| 17 | 25 | CONNECTOR |
| 18 | 25 | MICROPACKAGE |
| 19 | 50 | BOARD |
| 20 | 25 | CONNECTOR |
| 21 | 25 | MICROPACKAGE |
| S1 | 25 | MICROPACKAGE |
| S2 | 25 | CONNECTOR |
| S3 | 50 | BOARD |
| S4 | 625 | RIBBON CABLE |
| S5 | 625 | BOARD |
| S6 | 25 | CONNECTOR |
| S7 | 25 | MICROPACKAGE |

FIGURE 14

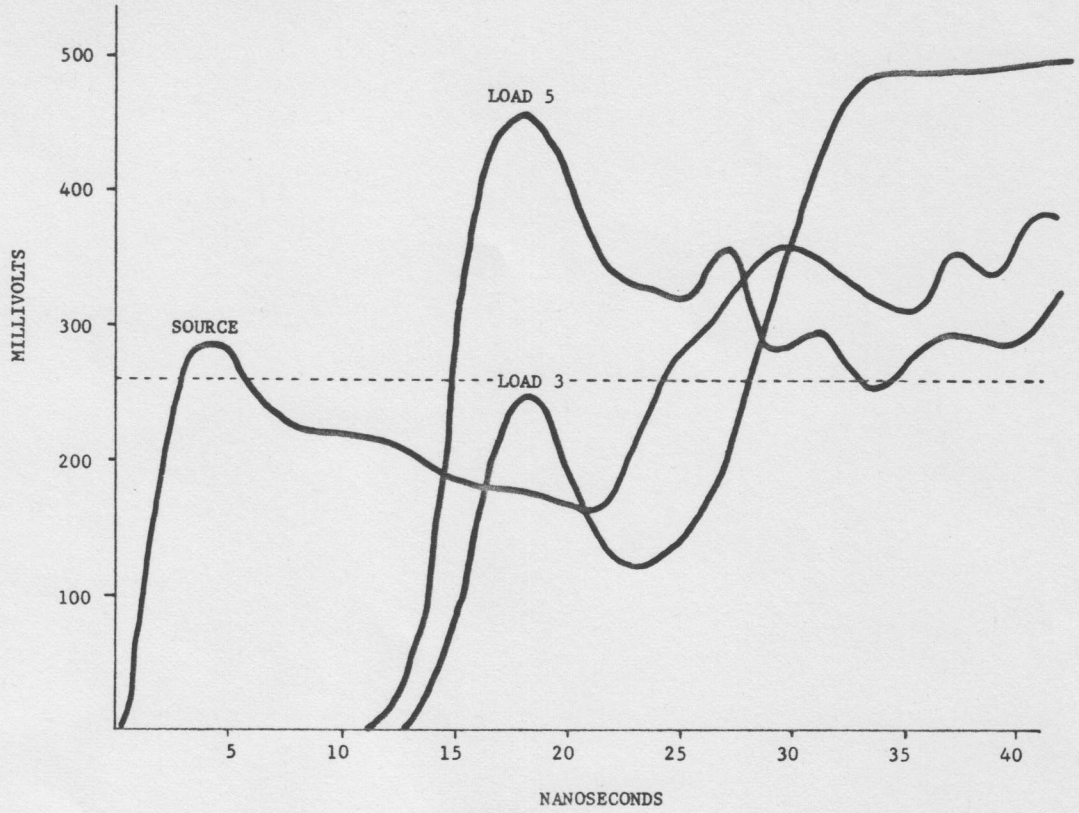



FIGURE 13

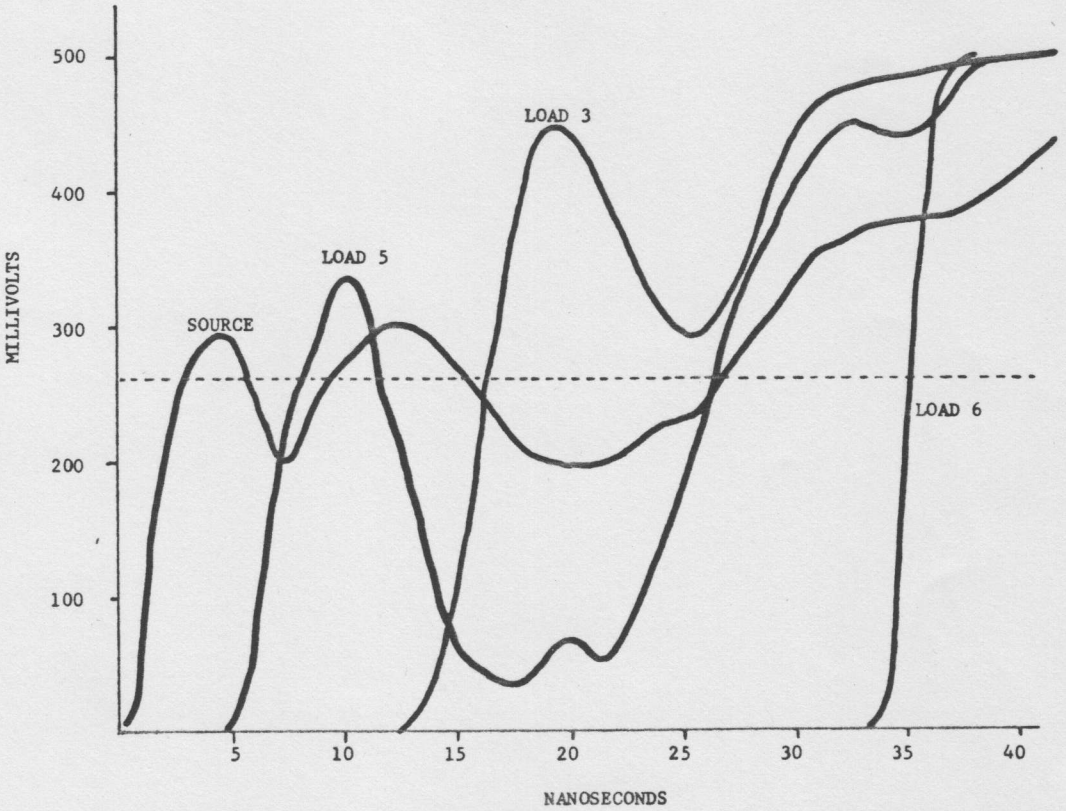



FIGURE 15

## APPENDIX - SIMULATION MODEL

The simulation program uses two separate models which interact in time. The two models are the transmission line (interconnect) model and a gate (load) model. By using the substructuring approach, the execution time and program storage requirements are kept to a minimum.

As is usually done, a segment of line may be broken into several elements. The basic current voltage relations for the line element are:

$$v = iR \quad , \quad i = kv \tag{A1}$$

$$v = L\frac{di}{dt} \quad , \quad i = m\int_o^t v\,dt \tag{A2}$$

$$v = S\int_o^t i dt \quad , \quad i = c\frac{dv}{dt} \tag{A3}$$

Using the second relations of equations (1), (2) and (3), the matrix equation for any line element may be written as:

$$\left([K] + \int_o^t [M]dt + \frac{d}{dt}[C]\right)\{V\} = \{I\} \tag{A4}$$

where the matrices are developed from a finite element formulation.

$$[K] = \begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \tag{A5}$$

$$[M] = \begin{bmatrix} m & -m \\ -m & m \end{bmatrix} \tag{A6}$$

$$[C] = \frac{1}{6}\begin{bmatrix} 2c & c \\ c & 2c \end{bmatrix} \tag{A7}$$

The elemental matrix equation (4) is directly assemblable into a total interconnect matrix and by using line elements instead of current loops (i.e., writing mesh equations) the logic required for generating a total interconnect from many segments with many branch points is simple and straightforward.

Equation (4) is also the basic form of the total interconnect equation (except that $[K]$, $[M]$, and $[C]$ now represent the assemblage of all elemental matrices into the total matrix). The first step toward "computerizing" equation (4) is to use discrete time steps.

This yields the form

$$\left([K] + [M]\Delta t + [C]/\Delta t\right)\{V\}_i =$$

$$\{I\}_i + [C]/\Delta t\,\{V\}_{i-1} - \sum_{j=1}^{i-1}\left([M]\{V\}_j\right)\Delta t \tag{A8}$$

where

$\Delta t$ = time increment

$i$ = subscript denoting time T

For the remainder of this development, the term $\Delta t$ will be absorbed into the matrices $[M]$ and $[C]$ for the sake of simplified equation writing.

The accuracy as well as the execution time will of course depend on the size of the time step. Using a constant time step allows doing only one matrix inversion and then just changing (i.e., updating) the right hand side of equation (8). A simple manipulation which allows larger time steps for the same accuracy (the amount of "larger" depends on the user's problem but can be on the order of a factor of two) is to write the equation for the voltage at time $T - \frac{1}{2}\Delta t$ instead of time T.

This means

$$\{V\}_{i-\frac{1}{2}} = \tfrac{1}{2}\left(\{V\}_i + \{V\}_{i-1}\right) \tag{A9}$$

in terms of subscripts. Applying equation (9) to equation (8) yields

$$\left(\tfrac{1}{2}[K] + 1/8\,[M] + [C]\right)\{V\}_i = ,$$

$$\{I\}_{i-\frac{1}{2}} - \tfrac{1}{2}[K]\{V\}_{i-1} + [C]\{V\}_{i-1}$$

$$-7/8\,[M]\{V\}_{i-1} - \sum_{j=1}^{i-2}\left([M]\{V\}_j\right) \tag{A10}$$

where the factors 1/8 and 7/8 come from using

$$\{V\}_{i-3/4} = \tfrac{1}{4}\{V\}_i = 3/4\,\{V\}_{i-1} \tag{A11}$$

for the integration in the interval from $T - \Delta t$ to $T - \Delta t/2$. Equation (10) will then be in the form

$$[D]\{V\}_i = \{I\}_{i-\frac{1}{2}} + f\left(\{V\}_{i-1}, \cdots, \{V\}_1\right) \tag{A12}$$

The input to the line model is a current generator with an idealized waveform which matches the rise time for actual experimental data.

By making use of symmetry and the fact that the matrix is primarily tri-diagonal except at branch nodes, the storage can be significantly reduced. For large problems, the storage technique used in this program can reduce the storage requirements by a factor of 100. Yet, the simultaneous equation solution used is direct (by matrix manipulation) and not iterative.

By not including the load gate model in the line model, the symmetry is preserved and the matrix storage is kept small. Also, by using a separate gate model, the voltage (hence time) dependent gate matrix is allocated separate storage and the need to modify matrix elements as a function of time is eliminated from the line matrix. This is important since matrix inversion being done only once (at time zero) is preserved with the accompanying order of magnitude execution time reduction as compared to inversion for each time step.

The gate model uses a charge control model for the transistors and the schematic is shown in Figure A1. The node numbers are specifically chosen to reduce the number of arithmetic operations required for solving the simultaneous equations. The parameters for the transistors are experimentally determined from actual gates. Writing the model equations will lead to the matrix form

$$[A]\{V\} = \{r\} \tag{A13}$$

where the right hand side ($\{r\}$) will be composed of a current vector, a vector containing capacitance terms multiplied by nodal voltages, and a vector of known voltages ($V_{in}$, $V_{26}$, $V_G$, $V_{33}$). The transistor capacitors are a function of the voltages across them, so they are calculated by using the nodal voltages from the preceeding time step. The current generators are very dependent on an exponential function of the spanning voltage and a better than constant (during the time step) current approximation is easily achieved. For example, in terms of

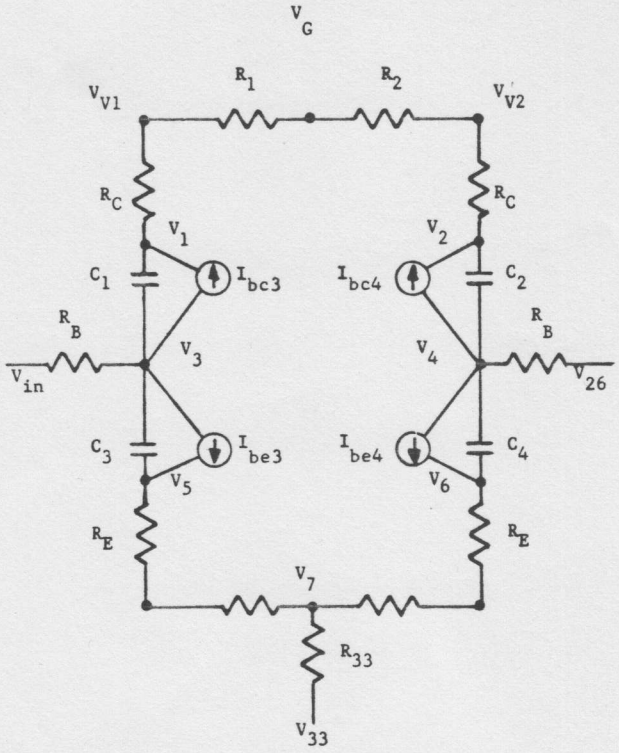



FIGURE A1

the voltage at the beginning of the time step ($V'_{bc}$), the current at the end of the time step may be written as

$$I_{bc} \approx -I_{cs} + I_{cs}\left(e^{\theta_I V'_{bc}} + e^{\theta_I V'_{bc}} (V_{bc} - V'_{bc})\,\theta_I\right). \qquad \text{(A14)}$$

This is obtained from the first two terms of a series expansion of the exponential factor. A similar expression also exists for $I_{be}$.

These expressions for current provide an implicit solution for the unknown voltages in terms of the current by contributing terms to the matrix [A] in (A13) which also destroys its symmetry. However, since the matrix is sparse, it can easily be inverted by hand and the solution can be coded directly into the program, further minimizing the execution time. This yields a directly computable set of equations for the elements of $[A^{-1}]$ which then provides the equations for

$$\{V\} = [A^{-1}]\{r\} \qquad \text{(A15)}$$

The subroutine for the gate model is used for all loads by having a set of vectors which contain the nodal voltages for each load and these are updated at each time step for use in the next time step.

The input voltage to the gate model is the voltage at time T on the line (interconnect) node to which a given load is attached. The nodal voltages in the gate model are solved for and the voltage on the internal end of the input resistor ($V_R$) is used along with the voltage from the preceeding time to predict the next half time step voltage, $V_R$. The loads are "attached" to the line model by modifying the [K] matrix in equation (A10) to include an input resistor at all load location nodes. The value of the input resistor is proportional to the load factor of the gate (i.e., "high" or "low" current gates). Then the predicted voltages ($V_R$'s) from the gate model are used as boundary conditions for the modified (A10) which now becomes (for T - $\Delta t/2$)

$$\left(\tfrac{1}{2}[K] + \tfrac{1}{2}[K]_R + 1/8\,[M] + [C]\right)\{V\}_i = \{I\}_{i-\frac{1}{2}} - \tfrac{1}{2}[K]\{V\}_{i-1} - \tfrac{1}{2}[K]_R \left(\{V\}_{i-1} - 3\{V_R\}_{i-1} + \{V_R\}_{i-2}\right) + [C]\{V\}_{i-1} - 7/8\,[M]\{V\}_{i-1} - \sum_{j=1}^{i-2}\left([M]\{V\}_j\right) \qquad \text{(A16)}$$